Natural Language Processing (CS 3730)

Lecture 1
August 27, 2001

Today's Outline

Course Overview

Administration

Natural Language Processing

- The field of *Natural Language Processing*, or *Computational Linguistics*, is primarily concerned with the creation of computer programs that perform useful and interesting tasks with human languages (e.g., understanding, generation, learning).
- It is secondarily concerned with using computational metaphors to help us come to a better understanding of human language.
- The foundations of the field are in computer science, linguistics, mathematics, electrical engineering, and psychology.

Major Topics of this Course

Knowledge of Language

- words (meaningful components of words)
- syntax (structure of a sentence)
- semantics (explicit meaning of a sentence)
- pragmatics/discourse (implicit/contextual meaning)

Methodologies

- knowledge-based and statistical
- state machines, rule systems, grammars, logic, search, probability, automata, machine learning, and more

Applications

- from hyphenators to intelligent agents

Knowledge of Language

- Example dialogue from *2001: A Space Odyssey*

```
Dave: Open the pod bay doors, HAL.
HAL: I'm sorry Dave, I'm afraid I can't do that.
```

- To participate in such a conversation, HAL needs knowledge about many levels of language
  - words: producing contractions, plurals
  - syntax: questions versus statements, word order and grouping
  - semantics: meaning of words in isolation and compositionally
  - pragmatics: politeness and indirectness
  - discourse: between utterance references

Applications

What makes a computer application a language processing application?

- language processing applications require the use of knowledge of language

Little Applications

Little applications typically make use of only a small amount of a single kind of knowledge of language. Many of these types of applications are currently in use, but they are often nearly invisible.

- line breakers
- hyphenators
- spelling correctors
- OCR software
- grammar and style checkers

Big Applications

Big applications often make use of large amounts and varied kinds of knowledge of language.

- information retrieval and extraction
- question answering
- dialogue management
- text summarization
- machine translation
- HAL

Applications, cont.

Output of state of the art programs presented in class

Administration

Professor

Textbook

Web page

Requirements

Who should be here

Dr. Diane Litman

Affiliations

- Associate Professor, Computer Science Department
- Research Scientist, LRDC

Contact Information

- 214 Mineral Industries Building, (412) 624-8838
- 741 LRDC
- litman@cs.pitt.edu
- http://www.cs.pitt.edu/~litman

Office Hours (tentative)

- M 1:30-2:30, W 10:30-11:30
- 214 Mineral Industries Building

Dr. Diane Litman, cont.

Education

- 1986: PhD, Computer Science, University of Rochester
- Thesis: Plan Recognition and Discourse Analysis: An Integrated Approach for Understanding Dialogues

Employment

- 1985-2001: Technical Staff, AI Principles Research Department., AT&T Labs - Research (formerly AT&T Bell Laboratories)
- 1990-1992: Assistant Professor, Computer Science, Columbia University

Research

- Natural Language Processing
  - discourse, dialogue, pragmatics, spoken language, learning
- Other Artificial Intelligence
  - knowledge representation, plan recognition, user modeling

Professional Activities

- 2001-2001: Chair, North American Chapter of the Association for Computational Linguistics
- Numerous program committees (see web page)

The Text

Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition
By Daniel Jurafsky and James H. Martin; Prentice-Hall, 2000

- http://www.cs.colorado.edu/∼martin/slp.html
- should be available from the bookstore
- also available from online providers

Course emphasis will be from a "computer science" perspective

Course Web Page

URL

- http://www.cs.pitt.edu/∼litman/courses/CS3730/cs3730.html

Syllabus

- topics
- readings
- assignments
- lecture notes
- announcements
- NOTE: involves viewing/printing postscript and pdf documents

Requirements

Readings (before class!)

Homeworks

- problem sets
- writing programs
- using programs

Exams

- midterm
- final

Class participation

Prerequisites

An interest in language, and at least some of the following:

- ability to write and use programs
- background in computer science, linguistics, or speech
- some basic exposure to logic, probability, algorithm analysis, and/or programming language theory would be helpful

For Next Time

Read Chapter 1 (available online)

Survey

Name:

Email:

Department:

Year:

Undergraduate Major:

Relevant Courses:

Relevant Research Experience:

Methodologies (state machines, rule systems, grammars, logic, search, probability, automata, machine learning, etc.):

Programming Languages:

Operating Systems:

Goals:

Acknowledgements

Some of this material was taken from the course web page of James Martin